第三章 最速下降算法

§3.1 最速下降算法的基本思想

- 考虑一个实数代价函数$J(\mathbf{w})$，它是某个向量$\mathbf{w}$ 的连续可导函数。
- 问题描述：寻找一个最优解 $\mathbf{w}_o$，使它满足

$$J(\mathbf{w}_o) \le J(\mathbf{w}), \quad \forall \mathbf{w}$$

- 在数值计算中，常用**迭代下降法**来解决这类问题。

- **迭代下降法**：从某一初始值$\mathbf{w}(0)$开始，按照固定的步骤，迭代产生一系列权值向量 $\mathbf{w}(1), \mathbf{w}(2), \mathbf{w}(3), \cdots$，使得代价函数的值在每一次迭代后都下降，即满足

$$J(\mathbf{w}(n+1)) < J(\mathbf{w}(n))$$

其中，$\mathbf{w}(n)$ 和 $\mathbf{w}(n+1)$分别是第n次迭代和第n+1次迭代得到的权值向量。

- 迭代下降的一种简单形式—**最速下降算法**：该方法是沿着最速下降方向(负梯度方向)连续调整权值向量$\mathbf{w}$。

- 将代价函数$J(\mathbf{w})$的梯度表示为

$$\boxed{\mathbf{g} = \nabla J(\mathbf{w})}$$

注意，此处删除了教材中的第二个等号及其之后的偏导数，教材中有误。

- 从而最速下降法可表示为(μ为步长)：

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \frac{1}{2}\mu\mathbf{g}(n)$$

- 从n次迭代到$n+1$次迭代，权值向量的调整量为

$$\delta \mathbf{w}(n) = \mathbf{w}(n+1) - \mathbf{w}(n) = -\frac{1}{2}\mu \mathbf{g}(n)$$

- **证明**：将$J(\mathbf{w}(n+1))$ 在$J(\mathbf{w}(n))$处进行一阶泰勒展开，可得

$$J(\mathbf{w}(n+1)) \approx J(\mathbf{w}(n)) + \mathbf{g}^H(n)\delta \mathbf{w}(n)$$

- 将权值向量的调整量$\delta \mathbf{w}(n)$代入上式，可得

$$J(\mathbf{w}(n+1)) \approx J(\mathbf{w}(n)) - \frac{1}{2}\|\mathbf{g}(n)\|^2$$

§3.2 最速下降算法应用于维纳滤波器

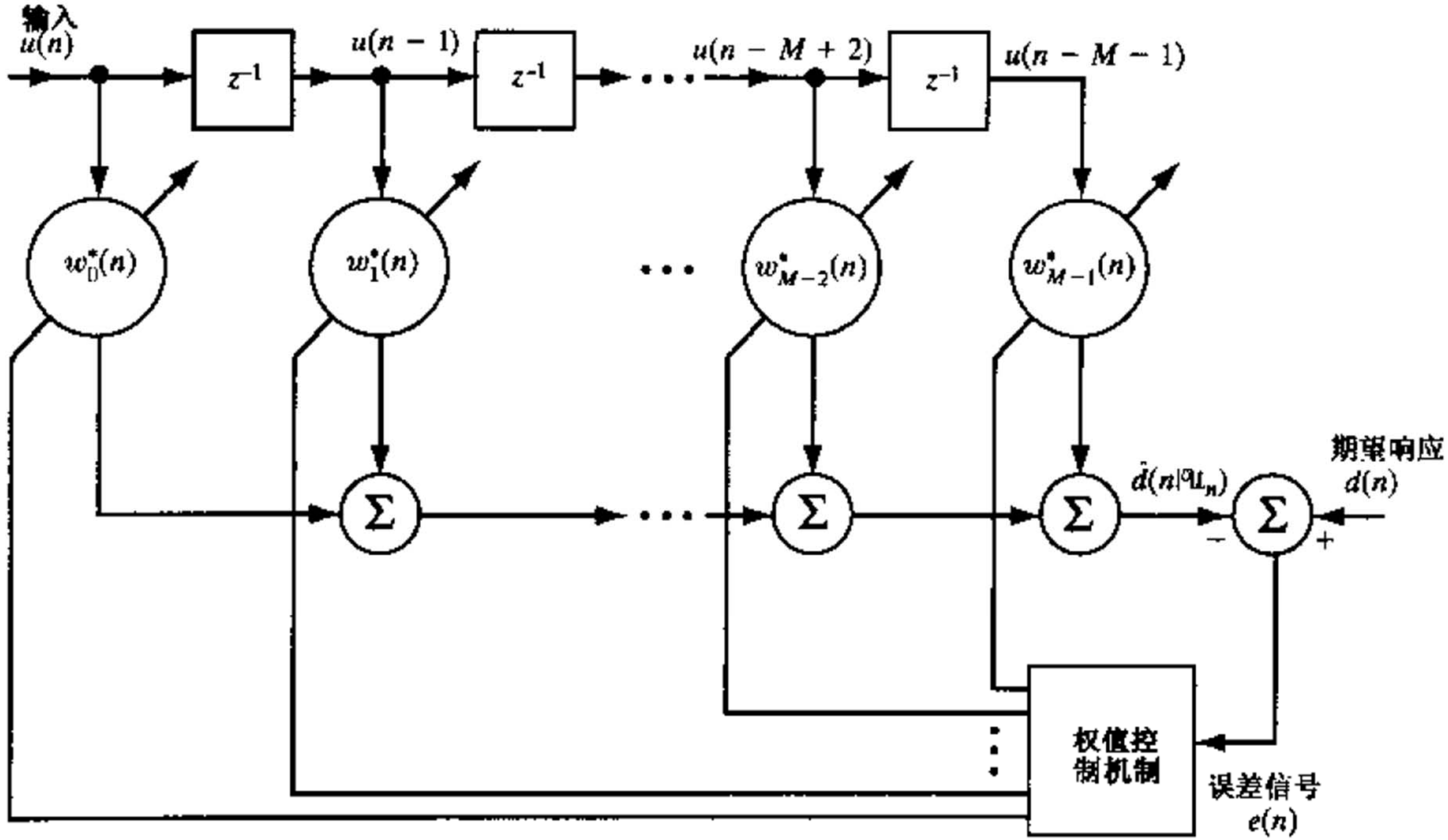

图 4.1　自适应横向滤波器的结构

- 横向滤波器的**抽头权值向量**和**输入向量分别**为

$$\mathbf{u}(n)=[u(n),u(n-1),\cdots,u(n-M+1)]^T$$

$$\mathbf{w}(n)=[w_0(n),w_1(n),\cdots,w_{M-1}(n)]^T$$

- 横向滤波器的输出相对于期望信号的误差为

$$e(n)=d(n)-\hat{d}(n\,|\,\mathcal{U}_n)=d(n)-\mathbf{w}^{\mathrm{H}}(n)\mathbf{u}(n)$$

- 第二章提到，维纳滤波器的代价函数为

$$J(n)=\mathrm{E}[|e(n)|^2]=\mathrm{E}[e(n)e^*(n)]$$

将 $e(n)=d(n)-\mathbf{w}^{\mathrm{H}}\mathbf{u}(n)$ 代入上式，可得

$$J(n)=\sigma_d^2-\mathbf{w}^{\mathrm{H}}(n)\mathbf{p}-\mathbf{p}^{\mathrm{H}}\mathbf{w}+\mathbf{w}^{\mathrm{H}}\mathbf{R}\mathbf{w}$$

其中：

$$\sigma_d^2=E[d^2(n)],$$

$$\mathbf{p}=E[\mathbf{u}(n)d^*(n)],$$

$$\mathbf{R}=E[\mathbf{u}(n)\mathbf{u}^H(n)]$$

- 由第二章可知，$J(n)$ 的梯度向量可写成

$$\nabla J(n)=\begin{bmatrix} \dfrac{\partial J(n)}{\partial a_0(n)}+\mathrm{j}\dfrac{\partial J(n)}{\partial b_0(n)} \\ \dfrac{\partial J(n)}{\partial a_1(n)}+\mathrm{j}\dfrac{\partial J(n)}{\partial b_1(n)} \\ \vdots \\ \dfrac{\partial J(n)}{\partial a_{M-1}(n)}+\mathrm{j}\dfrac{\partial J(n)}{\partial b_{M-1}(n)} \end{bmatrix}$$

$$=-2\mathbf{p}+2\mathbf{R}\mathbf{w}(n)$$

- 将 $\nabla J(n) = -2\mathbf{p} + 2\mathbf{R}\mathbf{w}$ 代入最速下降迭代公式 $\mathbf{w}(n+1) = \mathbf{w}(n) - (1/2)\mu\mathbf{g}(n)$ ，可得求维纳滤波器的迭代解

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu[\mathbf{p} - \mathbf{R}\mathbf{w}(n)]$$

- 一方面，与维纳滤波器的闭合解 $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$ 相比，迭代解不需要求相关矩阵 $\mathbf{R}$ 的逆
- 更重要的是，迭代解是经典的最小均方算法的基础。

§3.3 最速下降算法的稳定性

- 横向滤波器的**维纳解**为 $\mathbf{w}_o=\mathbf{R}^{-1}\mathbf{p}$

  其**迭代解**为 $\mathbf{w}(n+1)=\mathbf{w}(n)+\mu[\mathbf{p}-\mathbf{R}\mathbf{w}(n)]$
- **疑问**：当 $n\to\infty$ 时，是否有 $\mathbf{w}(n)\to\mathbf{w}_o$？

  若有，需要满足怎样的条件？
- 下面进行证明

- **证明：**
- 将n时刻的权值误差向量定义为

$$\mathbf{c}(n) = \mathbf{w}_o - \mathbf{w}(n)$$

- 将迭代式 $\mathbf{w}(n+1) = \mathbf{w}(n) + \mu[\mathbf{p} - \mathbf{R}\mathbf{w}(n)]$ 两边同减去 $\mathbf{w}_o$，可得

$$\mathbf{c}(n+1) = \mathbf{c}(n) - \mu[\mathbf{p} - \mathbf{R}\mathbf{w}(n)]$$

- 将维纳方程$\mathbf{R}\mathbf{w}_o = \mathbf{p}$代入上式，消去$\mathbf{p}$，可得

$$\mathbf{c}(n+1) = (\mathbf{I} - \mu\mathbf{R})\mathbf{c}(n)$$

- 使用特征值分解，可将相关矩阵表示为

$$\mathbf{R}=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{H}}$$

- 其中，$\mathbf{\Lambda}=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_M]$为**特征值矩阵**，$\mathbf{Q}$为变换的**酉矩阵**，满足$\mathbf{Q}^H\mathbf{Q}=\mathbf{I}$。则迭代式

$$\mathbf{c}(n+1)=(\mathbf{I}-\mu\mathbf{R})\mathbf{c}(n)$$

可变换为

$$\mathbf{c}(n+1)=(\mathbf{I}-\mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{H}})\mathbf{c}(n)$$

- 将式$\mathbf{c}(n+1)=(\mathbf{I}-\mu\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{H}})\mathbf{c}(n)$ 两边乘以酉矩阵的共轭转置$\mathbf{Q}^{\mathrm{H}}$，并利用性质$\mathbf{Q}^{H}\mathbf{Q}=\mathbf{I}$可得

$$\begin{aligned}\mathbf{Q}^{\mathrm{H}}\mathbf{c}(n+1)&=(\mathbf{Q}^{\mathrm{H}}-\mu\mathbf{\Lambda}\mathbf{Q}^{\mathrm{H}})\mathbf{c}(n)\\&=(\mathbf{I}-\mu\mathbf{\Lambda})\mathbf{Q}^{\mathrm{H}}\mathbf{c}(n)\end{aligned}$$

- 定义**变换向量** $\mathbf{v}(n)=\mathbf{Q}^{\mathrm{H}}\mathbf{c}(n)$，则有

$$\mathbf{v}(n+1)=(\mathbf{I}-\mu\mathbf{\Lambda})\mathbf{v}(n)$$

- 初始权值通常取为$\mathbf{w}(0)=\mathbf{0}$，则有

$$\mathbf{v}(0)=\mathbf{Q}^{\mathrm{H}}\mathbf{c}(0)=\mathbf{Q}^{\mathrm{H}}[\mathbf{w}_o-\mathbf{0}]=\mathbf{Q}^{\mathrm{H}}\mathbf{w}_o$$

- 若证明$n \to \infty$时，$\mathbf{w}(n) \to \mathbf{w}_o$

  可证明$n \to \infty$时，$\mathbf{c}(n) = \mathbf{w}_o - \mathbf{w}(n) \to \mathbf{0}$

  或证明$n \to \infty$时，$\mathbf{v}(n) = \mathbf{Q}^{\mathrm{H}}\mathbf{c}(n) \to \mathbf{0}$

- 令$v_k(n)$为$\mathbf{v}(n)$的第k个元素，$k = 1, 2, \cdots, M$ 。

  即要求**证明**：$n \to \infty$时，$v_k(n+1) \to 0$

- 式 $\mathbf{v}(n+1)=(\mathbf{I}-\mu\mathbf{\Lambda})\mathbf{v}(n)$ 的展开式为

$$\begin{bmatrix} v_1(n+1) \\ v_2(n+1) \\ \vdots \\ v_M(n+1) \end{bmatrix} = \begin{bmatrix} 1-\mu\lambda_1 & & & \\ & 1-\mu\lambda_2 & & \\ & & \ddots & \\ & & & 1-\mu\lambda_M \end{bmatrix} \begin{bmatrix} v_1(n) \\ v_2(n) \\ \vdots \\ v_M(n) \end{bmatrix}$$

- 或写成

$$\begin{aligned} v_1(n+1) &= (1-\mu\lambda_1)v_1(n) \\ v_2(n+1) &= (1-\mu\lambda_2)v_2(n) \\ &\vdots \\ v_M(n+1) &= (1-\mu\lambda_M)v_M(n) \end{aligned}$$

- 由第k个方程

$$v_k(n+1)=(1-\mu\lambda_k)v_k(n),\ k=1,2,\cdots,M$$

递推可得

$$v_k(n+1)=(1-\mu\lambda_k)^{n+1}v_k(0)$$

- 要使得 $v_k(\infty)\to 0$，上式中必须满足

$$-1<(1-\mu\lambda_k)<1,\ k=1,2,\cdots,M$$

- 因此，最速下降法的**稳定性条件**为 $0<\mu<\dfrac{2}{\lambda_{\max}}$，其中

$$\lambda_{\max}=\max\{\lambda_1,\lambda_1,\cdots,\lambda_M\}$$